Productivity and Quality in Health Care: Evidence from the Dialysis Industry*

Paul L.E. Grieco† Ryan C. McDevitt‡

March 29, 2014

Abstract

We show that healthcare providers face a tradeoff between increasing the number of patients they treat and improving their quality of care, with those providers facing the strongest incentives to treat more patients delivering the lowest quality of care. To measure the magnitude of this quality-quantity tradeoff, we estimate a model of dialysis provision that explicitly incorporates a center's endogenous choice of treatment quality and allows for unobserved differences in productivity across centers. We find that centers may treat 1 percent more patients by allowing their expected infection rate to increase by 0.8 percentage points (6 percent), holding inputs and productivity fixed. Our approach provides unbiased estimates of productivity, whereas traditional methods misattribute lower-quality care to greater productivity. We also find (i) extensive quality-adjusted productivity dispersion across providers, (ii) better outcomes among non-profit entities, and (iii) comparatively little effect from competition.

JEL: D24, I1, L2
Keywords: productivity; quality variation; health care

*We thank Matt Grennan, Darius Lakdawalla, David Rivers, Mark Roberts, and Frederic Warzynski for their helpful comments. We also thank participants at the 2012 International Industrial Organization Conference (Arlington, VA), the 2012 FTC Microeconomics Conference (Washington, DC), the 2013 Meetings of the Econometric Society (San Diego, CA), and the 2013 Annual Conference of the European Association for Research in Industrial Economics (Evora, Portugal). We'd also like to thank seminar participants at Columbia University, Drexel University, and the University of Toronto.

†The Pennsylvania State University, Department of Economics, paul.grieco@psu.edu
‡Duke University, The Fuqua School of Business, ryan.mcdevitt@duke.edu

1 Introduction

Rising healthcare expenditures have motivated spending reforms such as Medicare's prospective payment system, which ties reimbursements to a fixed amount per service irrespective of a provider's actual costs. While such initiatives aim to limit wasteful healthcare expenses, they may inadvertently result in lower-quality care: providers may face an incentive to reduce the quality of their treatments to minimize costs and increase patient loads. Our paper examines this tradeoff explicitly and provides policymakers with an empirical framework for measuring its magnitude within health care.

A prominent setting where such a tradeoff may be particularly acute and worthy of study is outpatient dialysis treatments, a process that cleans the blood of patients with end-stage renal disease (ESRD), or kidney failure. Payments to dialysis facilities comprise a substantial portion of Medicare's expenditures each year — over $20 billion in 2011, or 6% of total Medicare spending — and several features of the dialysis industry make it an appealing empirical setting to evaluate healthcare quality. First, payments for treatment are largely uniform due to Medicare's prospective payment system and do not depend on treatment quality, making it possible for us to isolate the effects of quality provision from price discrimination.[1] Second, dialysis treatments follow a straightforward process related to stations and staff, which allows us to approximate a facility's production function. Third, we observe centers' input levels (i.e., staffing and machines) and production (i.e., patient loads), which allows us to cleanly identify the relationship between inputs and outputs. Finally, facilities have observable differences in outcomes that relate directly to the quality of care they provide (e.g., infection and death rates), which allows us to connect a firm's inputs and outputs to its treatment quality, the primary aim of our research.

Determining whether dialysis centers do, in fact, face an incentive to trade off quality for quantity requires overcoming a key empirical challenge: providers' endogenous choices with respect to inputs and quality may bias estimates of the quality-quantity tradeoff. That is, because centers' input choices and targeted levels of quality are not exogenously assigned, estimating the relationship between quality and quantity becomes confounded by unobserved differences in pro-

[1] In 2012, Medicare instituted a Quality Incentive Program (QIP) for dialysis centers that reduces reimbursements by 2 percent if centers do not adhere to a quality standard for average hemoglobin levels and urea reductions rates, two measures of the effectiveness of dialysis treatment. However, although it is considered a novel attempt to incorporate quality standards into the Prospective Payment System, even the QIP does not account for infection rates — clearly an important measure of treatment quality — in its measurement system. The QIP was not in effect for the data that are currently available.

ductivity, such as managerial ability or patient characteristics.[2] As higher levels of productivity effectively shift out a center's production possibilities frontier, the center becomes able both to treat more patients and to provide better care; at the extreme, a positive correlation between quality and quantity may result. Even at modest levels of dispersion, this correlation will bias reduced-form estimates of the quality-quantity tradeoff and lead researchers to underestimate the true costs of improving treatment quality.

To uncover the cost of providing higher-quality care in a consistent manner, we build on the structural methods for estimating firm-level production functions first proposed by Olley & Pakes (1996), and later extended by Levinsohn & Petrin (2003), Ackerberg et al. (2006), Gandhi et al. (2011), and others. Conceptually, we adapt these methods to incorporate a "quality-choice" stage that comes after a firm's choices of labor and capital inputs. That is, after acquiring capital and training workers, a manager observes his center's expected level of productivity and chooses the quality of care to provide by, for example, stipulating guidelines for the length of treatment or cleanliness of equipment. Allowing for these endogenous quality choices is an important adaptation for healthcare settings because providers under a prospective payment system may appear more productive by treating many patients ineffectively, whereas policy makers have concerns over both productivity and effectiveness.

Because we do not directly observe firms' actual choices regarding quality, we instead use observable measures of patient outcomes as a proxy for what those choices must have been — if high-quality care is more likely to result in better health outcomes, those outcomes are valid proxies for quality choices. Using multiple measures of health outcomes (in our case, a center's septic infection and mortality rates), we can then implement an instrumental variable approach to recover the impact of quality choices on output.

We use our results to investigate why dialysis centers have such extensive variation in treatment quality, an important policy question for which our empirical approach offers unique insights. While differences in productivity could represent one source of the variation in quality, firms may also choose different quality-quantity combinations deliberately, even when they face the same production possibilities; notably, dialysis centers have an incentive to minimize the costs of treating patients under Medicare's prospective payment system, which may include providing low-quality — and hence, less costly — care. Counteracting this incentive, however,

[2]While we control for observable differences in patient characteristics, unobservable differences may still affect firms' input and quality choices.

are plausible motivations for providing high-quality treatments: centers must report quality statistics to Medicare which are then made public, and face intermittent inspections by state regulators (Ramanarayanan & Snyder 2011). In addition, patients have some choice over their dialysis providers, potentially leading centers to compete for patients by providing higher-quality care (Dai 2012). Finally, non-profit centers may have objectives for providing high-quality care unrelated to maximizing profits (Sloan 2000).

From our analysis, we find a substantial quality-quantity tradeoff for dialysis treatments: a center can increase its patient load by 1 percent by allowing a 0.8 percentage point higher septic infection rate, holding input levels and productivity constant; equivalently, holding the number of treated patients constant but allowing a one standard deviation increase in a center's expected infection rate decreases its costs by the equivalent of three full-time employees. In addition, our approach allows us to recover estimates of total factor productivity for each firm that properly account for endogenous quality choices, and we find substantial productivity dispersion across firms that is not explained by differences in treatment quality. Finally, we investigate the determinants of quality in the industry and find that for-profit dialysis centers provide significantly worse care, with an infection rate 1.5 percentage points (roughly 12 percent) higher than their non-profit counterparts. At the same time, local competition does not appear to lead centers to improve their treatment quality. Taken together, these results provide evidence that profit-based incentives to reduce costs may lead to lower-quality care and that competition has a limited impact on quality.

In addition to providing relevant policy analysis, this paper also contributes to the growing literature in empirical industrial organization on the estimation of production functions. These methods have a long history in economics, with many well-known econometric issues related to selection and simultaneity bias receiving considerable attention.[3] In light of this, recent work has developed structural techniques that use firms' observed input decisions to control for unobserved productivity shocks and overcome endogeneity problems.[4] We extend these methods to incorporate observable measures of output quality into the production function, which is necessary for healthcare applications. To our knowledge, we are the first to apply these methods to a healthcare setting with the goal of measuring a quality-quantity tradeoff.[5]

[3]See Syverson (2011) for a recent review.

[4]See, for example, Olley & Pakes (1996), Ackerberg et al. (2006), and Levinsohn & Petrin (2003).

[5]Romley & Goldman (2011) consider quality choices among hospitals using a revealed-preference approach rather than outcome-based quality measures. Lee et al. (2012) use a structural approach to measure the impact of healthcare IT on hospital productivity, but do not consider output quality.

The remainder of our paper continues in the following section with a description of the outpatient dialysis industry and our data sources. Section 3 develops our structural model for estimating a production function in the presence of an endogenous quality choices, while Section 4 outlines our methods for bringing the model to the data. Section 5 presents our estimation results. Finally, Section 6 concludes with a discussion of our findings' implications.

2 Empirical Setting and Data Description

The demand for dialysis treatments comes from patients afflicted with end-stage renal disease (ESRD), a chronic condition characterized by functional kidney failure that results in death if not treated properly. Patients with ESRD effectively have only two treatment options, a kidney transplant or dialysis. Due to the long wait-list for transplants, however, nearly all ESRD patients at some point must undergo dialysis, a process that cleans the blood of waste and excess fluids. Patients can receive different dialysis modalities, with hemodialysis, a method that circulates a patient's blood through a filtering device before returning it to the body, constituting 90.4 percent of treatments (Center for Medicare and Medicaid Services).

Patients receiving dialysis in the United States primarily do so at free-standing dialysis facilities, which collectively comprise over 90 percent of the market (USRDS 2010).[6] Medicare's ESRD program, instituted by an act of Congress in 1973, covers the majority of these patients; notably, all patients with ESRD become eligible for Medicare coverage, regardless of age, and the program now includes over 400,000 individuals. Today, Medicare spends more than $20 billion a year on dialysis care — about $77,000 per patient annually — which constitutes more than six percent of all Medicare spending despite affecting fewer than one percent of Medicare patients (ProPublica 2011). Beginning in 1983, Medicare has paid dialysis providers a fixed, prospective payment — the "composite rate" — for each outpatient treatment delivered, up to a maximum of three sessions per week per patient. Initially, the payment rate did not adjust for quality, length of treatment, dialysis dose, or patient characteristics, though Medicare began to adjust payments based on patient characteristics in 2005.

Dialysis treatments require constant supervision by trained medical professionals, as patients must remain connected to a station for 2-5 hours to filter impurities and remove excess fluid from their blood. Prior to treatment, staff connect the machine to a patient by inserting two

[6] Other options for receiving dialysis include hospital emergency rooms and in-home treatments.

lines into a vascular access and assess his condition. During treatment, staff must continually monitor patients to evaluate conditions (e.g., blood pressure) and to treat symptoms that arise (e.g., hypotension). Following treatment, staff disconnect a patient from the station and assess his condition a final time before discharge. As a result of this hands-on care, the cost per patient treated necessarily increases with the average duration of the treatments. Labor costs, which consist largely of nurses and technicians' wages, reflect this, accounting for approximately 70-75 percent of a facility's total variable costs (Ford & Kaserman 2000).

Centers employ different types of labor, with registered nurses (RNs) constituting the majority of staff. Technicians, who have less-extensive training than RNs, also treat patients but can do so with only a high-school diploma and in-house training (though they must eventually pass a state or national certification test). Notably, centers cannot quickly react to changes in productivity by hiring more workers due to training and certification requirements. Centers also must have board-certified physicians as medical directors, though often have no physician on site. Medicare does not mandate a specific staffing ratio for dialysis centers, although some states do.

Another significant decision for dialysis facilities is the number of stations to have in operation. Centers vary widely in terms of size, ranging from 1 to 80 stations. Based on industry reports, a typical dialysis station costs $16,000 and has a useful life of approximately seven years (Imerman & Otto 2004).

In addition to labor and capital decisions, firms also choose how much effort to put towards providing high-quality care. For example, dialysis sessions require up to one hour of preparation and cleaning, which can be shortened according to a manager's discretion and can directly affect treatment outcomes. Importantly, patients undergoing dialysis face a high risk of septic infection due to the exposure of their blood during treatment, with the risk depending on the cleanliness of the dialysis center. The center likely has considerable control over its targeted infection rate, as health professionals who follow straightforward procedures can effectively minimize their patients' risk of contracting infections (Pronovost et al. 2006). The decision to do so, however, comes with the tradeoff of treating fewer patients due to the capacity constraints of the facility, which will ultimately reduce the center's profits.

Because a facility's reimbursement per treatment does not vary with its duration under Medicare's prospective payment system, a facility's profit per treatment decreases as treatment times — and, hence, labor costs — increase. At the same time, the effectiveness and safety

of dialysis increases with its duration; for instance, longer treatment cycles have been linked to lower mortality rates (ProPublica 2011). Centers thus face a tradeoff between improving treatment quality and decreasing costs.[7] And though the costs of providing high-quality care are relatively clear, the benefits for dialysis centers are less straightforward. First, demand-side incentives appear weak because dialysis provides life-sustaining functions for patients, making their demand for treatments inelastic. Second, patients typically have few dialysis centers to choose from in any given market — the mean market share across the United States is 0.457 — and, since ESRD immobilizes those affected by it, travel costs limit market choice.[8] Finally, as discussed above, Medicare's payment system provides no direct financial incentive for providing high-quality care.

At the same time, firms may still have several possible incentives for delivering high-quality care. For instance, a facility that provides inadequate treatments may face increased regulatory scrutiny that further drives patients to competitors or results in decertification. Moreover, when a facility does face competition for patients, providing low-quality care may lead its patients to defect to other facilities that provide better care. Finally, some centers, particularly non-profit entities, may have motives to provide high-quality care beyond just profitability.

Data Sources We use several sources of data for our analysis. Our primary dataset comes from the Centers for Medicare and Medicaid Services (CMS) which contracts with the University of Michigan's Kidney Epidemiology and Cost Center to compile customized reports for each dialysis facility in the country. In December 2010, ProPublica, a non-profit organization dedicated to investigative journalism, obtained these reports under the Freedom of Information Act and posted them online. We systematically downloaded all individual reports covering 2004 — 2008 and constructed a usable dataset. The data include detailed center-level information on aggregated patient (e.g., age, gender, co-morbid conditions, etc.) and facility (e.g., number of stations and nurses, years in operation, etc.) characteristics.

Table 1 presents selected summary statistics from the data, and several variables deserve note. First, Medicare analyzes individual patient records and calculates the number of patient-years each dialysis center treats (e.g., a patient treated at a center for six months is accounted

[7] Critics allege that facilities may their sacrifice quality of care in pursuit of efficiency, turning over three to four shifts of patients a day. And while policy makers contend that technicians should not monitor more than four patients at once, patient-to-staff ratios exceed this guideline in many facilities. At the extreme, inspection reports allege that some clinics have allowed patients to soil themselves rather than interrupt dialysis (ProPublica 2011).

[8] We use Hospital Service Areas (HSA) as the market definition for this calculation.

Table 1: Summary Statistics.

Variable	Mean	St. Dev.
Patient Years	50.856	31.913
FTE Staff	13.496	7.933
Net Hiring	1.064	0.552
Zero Net Hiring	0.127	0.333
Stations	18.612	7.877
Septic Infection Rate	12.504	6.399
Death Rate Ratio	1.041	0.405
Number of Firms	4,270	
Number of Firm-Years	18,295	

for as one half of a patient-year). We use this variable as our measure of output, as it provides an accurate record of dialysis provision that accounts for partial years of service due to death, transfers, transplants, newly diagnosed patients, and so forth. We also use the number of full-time equivalent (a weighted mix of full-time and part-time) employees at each center and the number of dialysis stations as our measures of labor and capital inputs, respectively. In terms of capital stock, the average number of dialysis stations used by a center is 18, making the purchase of a new machine a significant investment; reflecting this, firms have zero net investment for 90 percent of the center-year observations in the data. In terms of hiring, centers, on average, increase their staff by the equivalent of one full-time employee each year, while 12.7 percent of centers have no net change in employment in a given year.

We use a center's hospitalization rate from septic (blood) infections as our primary measure of quality, which averages 12.5 percent per year and has a standard deviation of over 6 percent. In addition to the septic infection rate, we use the ratio of deaths to expected deaths as an alternative measure of quality.[9] Importantly, we can also control for aggregate patient characteristics at each center that influence productivity and quality, which we discuss at length in Section 4.

The competitive environment faced by dialysis centers is highly variable, as shown in Table 2. Following the healthcare literature, we use hospital service areas (HSA) as our market definition for dialysis centers. While roughly 26 percent of dialysis centers are monopolies within their HSA, the average number of centers in an area is 8.1; in addition, the mean patient-weighted market share across centers within an HSA is 0.457.

[9]The center-level expected death rate is calculated by Medicare using individual patient characteristics.

Table 2: Number of Competitors within a Firm's HSA.

Num Comp.	N	Freq.	Cum.
0	4,789	0.2618	0.2618
1	3,223	0.1762	0.4379
2	1,828	0.0999	0.5379
3+	8,455	0.4621	1.0000

3 A Model of the Quality-Quantity Tradeoff in Dialysis

To measure the relationship between a firm's productivity and its treatment quality, we propose and estimate a structural model of dialysis provision. In doing so, we account for both the standard endogeneity problems associated with using observed input choices to estimate production functions and the additional problem introduced by a firm's endogenous choice of treatment quality. The complication related to endogenous quality decisions stems from the unobserved (to the econometrician) choice made by firms that receive positive shocks to productivity: they may choose either to treat more patients, or to treat current patients more intensively. If highly productive firms choose to provide higher-quality care for their patients, naïve estimates of the quality-quantity tradeoff will be biased, leading us to underestimate the true cost of delivering high-quality care.

To control for this potential source of bias, we extend the work of Olley & Pakes (1996) and Ackerberg et al. (2006) by incorporating firms' endogenous quality targets. And because we only observe noisy measures of quality in our data, we also control for measurement error in quality choices, proxied for by firm-level hospitalization rates for septic infection in our application. Specifically, the attenuation bias introduced by measurement error in quality choices would cause us to underestimate the magnitude of the quality-quantity tradeoff, which we correct for using an IV approach.

3.1 The Production Technology

We model the provision of dialysis treatments as a stochastic two-output production process, where the outputs are the number of patients treated and the quality of treatment provided. Conditional on inputs and productivity, firms form beliefs as to what combinations of output and quality they can achieve. They are aware of a tradeoff between quality and output that takes the form of a production possibilities frontier relating the expected number of patients

they treat and our measure of quality, the expected infection rate. Formally,

$$T(\tilde{y}_{it}, \tilde{q}_{it}) = F(k_{it}, \ell_{it}, \omega_{it}), \tag{1}$$

where $F(\cdot)$ is the production function with (log) capital, k_{it}, and (log) labor, ℓ_{it}, inputs, as well as firms' unobserved assessment of its own productivity, ω_{it}. We use the number of stations as our measure of capital and the full-time equivalent number of nurses and technicians as our measure of labor. The unobserved productivity term, ω_{it}, is intended to account for all factors observable to the firm but not to the econometrician that impact its production possibilities, such as the center's square footage, managerial ability, labor or capital quality, or patient characteristics; this last source of unobserved productivity is particularly important in a healthcare setting such as dialysis where patient sorting may induce large differences in each center's ability to treat patients. For example, highly educated patients may follow treatment protocols more closely and therefore require less attention from technicians while being treated. Although our data will allow us to control for a number of key patient characteristics, some will remain unobservable and must be captured by ω_{it}.

The transformation function, $T(\cdot)$, determines how the center's productive capacity is divided between each expected output. The first output, $\tilde{y}_{it}$, is the (log) expected number of patients treated by the center. The second, $\tilde{q}_{it}$, represents the expected quality of the treatments, which we model as a scalar index. In general, "quality" can have many dimensions for patients, such as the probability of becoming sick, the amount of time spent waiting for treatments, the convenience of the center's operating hours, or even having televisions available during treatments. Despite this, we focus on one specific dimension of quality, the probability that the patient will contract a septic infection due to treatment, arguably the most prominent dimension of quality due to its severe impact on patients' well-being. Septic infections occur among dialysis patients because their blood is exposed to the dialysis machine for an extended period of time and multiple patients use the same machine sequentially. Thoroughly cleaning dialysis machines reduces the risk of patients contracting an infection, but is costly because employees must devote time to the process and the machine cannot be used for treatments while being cleaned. Therefore, improving quality (i.e., reducing the firm's expected infection rate) requires the center to treat fewer patients, holding all else fixed. Uncovering this relationship between quantity and quality is the primary goal of our analysis.

3.2 The Timing of Dialysis Center Decision Making

In their seminal paper, Olley & Pakes (1996) use capital investment as a proxy for unobserved productivity under the motivation that firms with greater productivity, all else equal, will make larger investments. Given this intuition, differences in investments will provide a meaningful indication of differences in productivity. While natural for their setting of telecommunications equipment, this approach is not appropriate for dialysis centers because investment in new stations is too infrequent: investment is zero for over 90 percent of the firm-year observations in the data. In light of this, we instead use firms' hiring decisions, which provide a more natural proxy in our setting. Nurses and technicians employed by dialysis centers require training and credentialing, which introduce costs and time lags to hiring and layoff decisions. Therefore, we regard labor as a dynamic variable, which allows us to use a firm's (net) hiring decision to recover ω_{it}.[10]

In contrast to labor choices, a firm can quickly adjust the quality of care it provides. For example, to improve quality, a manager could advise his center's staff to take extra precautions when treating patients, or to reduce quality by placing less emphasis on cleanliness and more on speed (Pronovost et al. 2006). While a center can dictate these policy changes more quickly than it can make hiring or investment changes, a lag still exists between a manager's quality decision and its actual implementation.

A firm's manager makes these investment, hiring, and quality choices based on his center's capital stock, labor productivity, and a vector of other observable characteristics, x_{it}. Importantly, the components of x_{it} may affect the firm's policy function even though they do not affect production directly, and may include the extent of competition in the market, the firm's taste for quality via it's non-profit status, and other related variables. This leads to the timing assumptions of our model:

1. *Quality choice.* Firms begin the period knowing their current levels of capital k_{it} and labor ℓ_{it}, as well as a vector of observable state variables x_{it} which affect the firms preferences but not it's productive capacity. (For example, it observes whether it is for profit or non-profit, and the characteristics of the surrounding market.).[11] It also observes $\omega_{it}^q = E[\omega_{it}|I_{it}]$. It's expectation of this period's productivity given its start of period information set. With

[10] Note that this assumption conflicts with OP's conception of labor representing an immediately flexible input, though the distinction fits our setting.

[11] Note that these features may be correlated with unobserved productivity. In other words, the model allows that non-profits are less productive than for profits due to their having lower ω^q.

this information the center chooses a pair $(\tilde{y}_{it}, \tilde{q}_{it})$, the expected level of output and quality for the period.

2. *Production Occurs.* Based on it's chosen target, the center treats patients and observes outcomes. It learns the true number of patients treated and number of infections, which are reported in the data. The firm also learns it's productivity for the period, ω_{it} as well as two idiosyncratic shocks to its output and quality outcomes.

$$y_{it} = \tilde{y}_{it} + \varepsilon^y_{it}$$
$$q_{it} = \tilde{q}_{it} + \varepsilon^q_{it}$$
$$\omega^h_{it} = \omega^q_{it} + \varepsilon^\omega_{it}$$

Note that $\varepsilon_{it} = (\varepsilon^y_{it}, \varepsilon^q_{it}, \varepsilon^\omega_{it})$ are uncorrelated with the information available during the quality choice by construction. However they may be correlated with each other. That is, conditional on both ε^y_{it} and ε^q_{it} being positive, we would expect the firm to raise its assessment of its own productivity. However, it is not necessary for us to explicitly model the learning process of the firm.[12]

3. *Hiring and Investment Choice.* After observing production, the firm's state is updated to reflect what has been learned about its productivity, so the state is now $(k_{it}, \ell_{it}, x_{it}, \omega^h_{it})$. With this information, the firm decides on hiring and investment for the following period. Newly hired workers (and newly invested capital) do not become available until period $t+1$, making the transitions for labor and capital:

$$k_{i,t+1} = k_{i,t} + i_{i,t} \qquad \ell_{i,t+1} = \ell_{i,t} + h_{i,t}.$$

Our decision to model hiring with a lag reelects the fact that employment credentials and other adjustment costs are significant in the dialysis industry relative to the difficulty of altering workers on-the-job incentives to strive for either high output or high quality outcomes through the choice of $\tilde{q}$.

4. *New State Realized.* In line with the literature, we assume productivity expectations follow

[12]Without loss of generality, we could allow productivity within the period to evolve according to an unknown stochastically increasing Markov process. Letting it evolve according to a random walk is notationally convenient because $E[\omega^h|\omega^q] = \omega^q$.

an exogenous Markov process between periods t, and $t+1$:

$$E[\omega^q_{i,t+1}|I_{i,t}] = E[\omega^q_{i,t}|\omega_{i,t-1}]$$

where I_{it} represents firm i's information set at the end of peirod t. Also following the literature we assume this process is stochastically increasing in $\omega^h_{i,t-1}$ (Pakes 1994). We also assume that the state variable x_{it} also moves according to an exogenous Markov process (similar to De Loecker 2011).

In this setting, unobserved productivity encompasses any factor that allows a center to treat more patients given its observable characteristics and quality target. For instance, a center's patients may follow treatment protocols more closely than other centers' patients do, which then frees the center either (i) to treat more patients because it devotes less time to dealing with complications that arise, or (ii) to spend additional time treating existing patients more intensively, which ultimately improves outcomes but does not appear in raw productivity measures, such as output-to-labor ratios.

3.3 The Center's Quality Choice Problem

The center enters the quality choice stage of a period with a state variable (k, ℓ, x, ω^q). Based on its expectations of its productive capacity, it chooses $(\tilde{y}, \tilde{q})$, its targeted level of output and quality for the period. We assume that demand for dialysis is inelastically supplied, which is reasonable given the capacity restrictions in the industry (which we model though the production frontier) and waitlists for treatment in many markets. However, given its inputs, the center faces a production possibilities frontier that determines how many patients it can treat at a given quality level. Since the center will learn more about it's productivity in production, it optimizes it's quality choice under uncertainty. Since we assume that the quality choice is fully flexible and that the quality and output outcomes do not affect future states, the center's quality choice problem does not have dynamic considerations, the center chooses its expected quality

and output level to solve the static problem,

$$\begin{aligned} \pi(k,\ell,x,\omega^q) = \max_{\tilde{y},\tilde{q}} \; & E[\rho(y,q,k,\ell,x)] \\ \text{subject to: } & T(\tilde{y},\tilde{q}) \leq F(k,\ell,\omega^q) \\ & y = \tilde{y} + \varepsilon^y \\ & q = \tilde{q} + \varepsilon^q \end{aligned} \tag{2}$$

Here, $\rho(\cdot)$ represents the centers period return for output and quality given the current state. Dialysis center's objectives are difficult to model directly, so we are agnostic as to the precise form of this function. Most clearly, not all centers are profit maximizing. Moreover, even for-profit centers may see value in treating as many patients as effectively as possible, rather than taking a more narrow view of their objective as maximizing period profits. Moreover, to profit from treating patients, centers face potential tort litigation or additional oversight if patients outcomes are too low, however exactly how to build these into firm objectives is unclear, but is clearly related to their quality of care.

We assume that period payoffs incorporating these considerations can be summarized by $\rho(\cdot)$ which is increasing in the center's two outcomes, y (output) and q (quality). The state vector may play a critical role in altering how centers view payoffs. For example, given the structure of the prospective payment system, one might expect for profit centers to place a higher priority on y relative to q than non-profits. Alternatively, if a center is aware that it will be inspected this period, it may be more concerned with quality then a center that will not undergo inspection. Allowing $\rho(\cdot)$ to depend on profit status, enables for-profits and non-profits to have different quality policies even though they share the same production technology conditional on the rest of the state. This variation in firm policies is key to our estimation strategy, since in order to estimate the quality-quantity tradeoff we need to observe firms with similar production possibilities who choose different quality-quantity pairs.

The assumption that the number of patients served today does not affect the state of the firm tomorrow is extremely common. The assumption that the implications of quality are static is more strong, due to the possibility of reputation effects. However one could imagine accounting for reputation effects through period profits (e.g., the firm pays for low quality performance immediately). Extending the model to allow for a long run reputation would require an additional state variable and a precise model of how quality affects reputation.

Examining the centers' quality problem, the fact that $\rho(\cdot)$ are increasing in q and y guarantees that the constraint will bind, so there is a one-to-one mapping between $\tilde{q}$ and $\tilde{y}$ through the production constraint. Moreover, the following lemma establishes that the return to labor is increasing in productivity, which is important to establishing Proposition 1 below.

Lemma 1. *The center's expected period-return to labor is increasing in ω^q. I.e., $\frac{\partial \pi}{\partial \ell}$ is increasing in ω^q.*

The proof of this theorem is provided in the appendix. Intuitively, both increases in ℓ and ω^q relax the production constraint, which must always bind if the firm is acting optimally, due to non-satiation. The fact that the constraint binds implies that the return to increasing ℓ is increasing in any variable whose effect is only to relaxes the constraint further, such as ω^q.

3.4 The Center's Hiring and Investment Problem

After production, the center chooses hiring and investment for the following period. The Bellman equation for this choice is,

$$V^h(k, \ell, x, \omega^h) = \max_{i,h} -c(i,h) + \beta E[V^q(k+i, \ell+h, x', \omega^{q'})|k, \ell, \omega^h, i, h] \tag{3}$$

The function $c(\cdot)$ represents investment and hiring costs, V^q represents the value of the firm at the start of the period,

$$V^q(k, \ell, x, \omega^q) = \pi(k, \ell, x, \omega^q) + E[V^h(k, \ell, x, \omega^h)|k, \ell, x, \omega^q].$$

This slightly cumbersome notation is needed to account for the fact that the center's perception of its own productivity evolves over the course of the period from ω^q to ω^h as a result of the center observing its own production process.

Based on the lumpiness of investment in the industry, we assume that the choice of next period capital is discrete. On the other hand, we view the hiring choice as effectively continuous. This seems reasonable given the number of nurses in the industry and the ability to adjust nurse's hours from period to period. Given these assumptions, the following proposition establishes that, for a given level of investment, the there is a one-to-one relationship between ω^h and the center's hiring choice, $h(k, \ell, x, \omega^h)$. This result is critical to our estimation strategy, which will

use investment and hiring decisions jointly to recover unobserved productivity.

Proposition 1. *For any fixed investment level ι, firm hiring function $h(k, \ell, x, \omega^h)$ is invertible with respect to ω^h on the domain $\{(k, \ell, x, \omega^h) : i(k, \ell, x, \omega^h) = \iota\}$,*

$$\omega^h = h_\iota^{-1}(k, \ell, x, h).$$

The proof of this theorem makes use of results in (Pakes 1994, Theorem 1) and (De Loecker 2011, Appendix C). We show that, given Lemma 1, our problem can be written in such a way that (Pakes 1994, Theorem 1) can be applied directly. There is an added complication that we must also control for the discrete investment level chosen by the firm. However, since firms do not invest 92% of the observed periods, this complication is mild.

4 Estimation

We use the model to estimate the underlying parameters of the production function and recover each firm's unobserved productivity in every period. We adopt the following parsimonious functional forms to describe the transformation and production functions,

$$T(\tilde{y}_{it}, \tilde{q}_{it}) = \tilde{y}_{it} + \alpha_q \tilde{q}_{it} \tag{4}$$

$$F(k_{it}, \ell_{it}, \omega_{it}^q) = \beta_k k_{it} + \beta_\ell \ell_{it} + \omega_{it}^q \tag{5}$$

In short, we follow the common practice in the literature of assuming a Cobb-Douglass production function, where ω_{it} is a Hicks-neutral technology shifter. For the transformation function, we also assume a Cobb-Douglass like specification that parameterizes the production possibilities frontier by assuming that reducing the infection rate 1 percentage point (i.e., increasing $\tilde{q}_{it}$ by 1) will reduce expected output by a factor of α_q, which is constant across firms.

This specification allows us to connect a firm's quality target to observable outcomes in a direct manner. By increasing the effort it puts towards providing high-quality treatments, the firm incurs additional costs but increases the probability of delivering better treatment outcomes — that is, the firm may treat fewer patients with the same level of inputs. On the other hand, a change in inputs or productivity shifts the production possibilities frontier, but does not alter the transformation between outputs. A center with healthier patients recognizes that its production

frontier has shifted outward, but still faces a tradeoff between treating more patients at a given level of quality and providing higher-quality care for a given number of treatments.

In the data, we do not observe firms' expected output and quality. Instead, we observe realized patient loads and infection rates, which are subject to both measurement error and unanticipated shocks. To account for this, we assume that observed output is $Y_{it} = \tilde{Y}_{it} e^{-\varepsilon^y_{it}}$ and the observed infection rate is $q_{it} = \tilde{q}_{it} + \varepsilon^q$. Substituting these into (1), we arrive at the linear equation,

$$y_{it} = -\alpha_q q_{it} + \beta_k k_{it} + \beta_\ell \ell_{it} + \omega^q_{it} - \alpha\varepsilon^q_{it} + \varepsilon^y_{it}, \tag{6}$$

Suppose we were to estimate (6) by ordinary least squares with data on (y, q, k, ℓ). Then, the composite error term is $\omega^q_{it} + \alpha\varepsilon^q_{it} + \varepsilon^y_{it}$, and two sources of bias are immediately apparent: one due to ω_{it}, and the other due to ε^q_{it}.

First, we have the well-known endogeneity problem associated with estimating production functions: because ω_{it} is observed by the firm but not the econometrician, it may be correlated with the firm's capital and labor choices. Our approach adds an additional endogeneity problem, as ω_{it} may also affect the firm's quality target. As a result, OLS estimates of (6) are inconsistent. Classical methods of correcting for endogeneity involve applying instruments for capital, labor, and quality, or assuming productivity is fixed over time (i.e., $\omega_{it} = \omega_i$) and using a fixed-effects estimator (Mundalk 1961). In application, these approaches have had limited success. While input prices would seem to be appropriate instruments for capital and labor choices, they often have weak predictive power and can be difficult to obtain. A valid instrument for quality targets that is uncorrelated with unobserved productivity would be even more challenging to find. Furthermore, while the fixed-effects assumption is relatively easy to implement, it is quite strong and would not resolve the endogeneity problems if changes in productivity are responsible for changes in input (or, in our case, quality) choices.

To address these issues in a manufacturing context, Olley & Pakes (1996) propose an explicit structural approach to estimate the production process which uses observed firm decisions as proxies for unobserved productivity shocks, with the basic ideas behind this method extended further by Levinsohn & Petrin (2003) and Ackerberg et al. (2006).[13] We extend this approach to a health care context. In this context, productivity differences may be due to unobserved differences in inputs and management practices, but also due to unobservable differences in

[13] A second approach to production function estimation comes from the dynamic panel literature (e.g., Blundell & Bond 2000); Ackerberg et al. (2006) provides a comparison of these approaches.

patient mixes which may make patients harder or easier to treat at a given level of quality.

A second source of bias results from the error term, ε_{it}^{q}. Although this error is unanticipated by the firm, it is, by definition, correlated with our proxy for treatment quality, the observed infection rate q_{it}. This form of classical measurement error will induce attenuation bias, moving our estimate of α_q towards zero. We will address this issue by instrumenting for q_{it} with a second proxy for treatment quality, the center's "unexpected" death rate. If the unobservable (to the researcher) factors that lead to infections are uncorrelated with those that cause death, then the instrument is valid and we can consistently estimate α_q.[14] In the event that they are correlated and our instrument is invalid, our estimate of α_q remains biased towards zero and is best understood as a lower bound, making our results conservative.

Estimation proceeds in three steps. First, because we do not observe quality directly, we must find an appropriate proxy for quality based on center-level outcomes. Second, we specify the observed policy shifters, x_{it}, which we include in the firm's hiring function. Finally, we adapt the standard two-stage estimation strategy to incorporate an endogenous quality choice with a noisy proxy.

4.1 Proxy for the Quality Target

Although we do not observe treatment quality directly, the data contain information on patient outcomes that are correlated with a center's choices on this dimension. In particular, we focus on the center's infection rate as an indicator of quality. This is only an imperfect measure, however, because variation in the infection rate may be due to differences in patient characteristics across centers rather than differences in centers' quality choices. To account for this, we control for center-level averages of several patient characteristics that influence infection rates. Specifically, we use the (negative) residual from a regression of infection rates on patient characteristics as our proxy for patient quality; this residual represents the variation in infection rates that remains after controlling for observable differences in the patient pool, and therefore serves as a proxy for the center's target for providing high-quality treatments.

We for control observable several patient characteristics that influence a center's infection

[14] It is possible that the unobservable factors related to contracting an infection are correlated with the center's death rate. Note, however, that the unobservable factors from the researcher's perspective are observable to the firm (e.g., a patient with AIDS is both more likely to contract an infection and to die) are accounted for in our model through ω_{it}, and not the unanticipated quality shock, ε_{it}^{q}, and so would not induce such correlation. Because we estimate a strong quality-quantity tradeoff, our results are robust to this potential confound.

Table 3: Patient Characteristics Summary Statistics.

Variable	Mean	St. Dev.
Avg. Patient Age	61.518	4.381
Pct. Female	45.798	8.333
Pct. AV Fistula	43.016	13.477
Avg. Comorbid Conditions	3.026	0.826
Avg. Duration of ESRD	4.089	0.953
Avg. Hemoglobin Level	11.882	0.332
Number of Firm-Years	18,295	

rate beyond its quality decision, with summary statistics displayed in Table 3. Most notably, we include controls for patients' methods of vascular access, which can be either an arteriovenous (AV) fistula, AV graft, or venous catheter. A patient's vascular access method influences his likelihood of developing a blood infection, with an AV fistula significantly less likely to cause clots or infections. Centers vary in the proportion of patients with an AV fistula, which ultimately may affect treatment outcomes. In addition to a patient's method of vascular access, other characteristics may directly affect treatment outcomes. Because centers vary in terms of their patients' characteristics, we also include controls for patients' (i) average number of comorbid conditions, (ii) average duration of ESRD, (iii) average age, (iv) gender distribution, and (v) average hemoglobin levels.[15] Putting these center-level average patient characteristics together into the vector z_{it}, we estimate,

$$f_{it} = z_{it}\gamma - q_{it},$$

where f_{it} is the realized infection rate at center i in period t. The residuals from this regression reflect the center's relative infection rate after controlling for observable patient characteristics, which we then use as our measure of center quality.

Of course, some unobservable differences in patient health may remain even controlling for observables. Some of these may be observable to firms making their quality choice. Within our model, these unobservable differences are interpreted as differences in ω_{it} across firms. Thus, we will rely on the control function approach below to control for these differences as well as other unobservable differences in productivity (e.g., management ability or unobserved quality of inputs).

[15] Low hemoglobin levels are associated with anemia and pose health risks for dialysis patients.

Finally, because our proxy is a measure of outcomes, rather than firm expectations, it is also subject to measurement error, leading to attenuation bias. We have accounted for expectational or measurement error in our specification of the production function by including ε_{it}^{q}. To control for measurement error, we employ a second outcome variable as an instrument for quality targets. Specifically, we use Medicare's estimates for each center's expected death rate which are based on individual patient characteristics (individual-level characteristics are not released to protect patient privacy). Medicare uses this ratio as an indicator of center quality in its own reports, and we use this measure as a second noisy proxy for a center's quality. Under the assumption that the measurement error components of our two proxies are uncorrelated, the expected death rate is a valid instrument to consistently estimate α_q using the infection rate residual as a proxy.

4.2 Controlling for Policy Shifters

To invert the hiring function and recover each firm's productivity, we must explicitly control for all factors that affect hiring other than productivity. In our specification, we include the following sources of variation in x.

For-profit Status Centers differ in their ownership type, with roughly 87.7 percent operating as for-profit entities and the remainder as non-profit. A center's ownership structure may affect its polices related to hiring and treatment quality, and we therefore control for this distinction by including a dummy variable for the center's for-profit status in x_{it}.

Competition Because demand for dialysis treatments is local, the extent of competition a center faces may affect its hiring and quality choices. For instance, centers in highly competitive markets may choose to improve quality or increase staff levels to attract patients. We include the level of competition each center faces in x_{it} in the form of dummy variables for having 0, 1, 2, or 3 or more competitors in an HSA. We assume that entry is exogenous and realized at the beginning of the period, so the firm observes its competitors when making its quality and hiring choices.

4.3 Two-Step Estimation

We now turn our method for recovering the parameters of the production frontier. We first note that $\omega_{it}^h = \omega_{it}^q + \varepsilon_{it}^\omega$, so we can rewrite (6) as,

$$y_{it} = -\alpha_q q_{it} + \beta_k k_{it} + \beta_\ell \ell_{it} + \omega_{it}^h - \varepsilon_{it}^\omega - \alpha\varepsilon_{it}^q + \varepsilon_{it}^y.$$

Because $(\varepsilon_{it}^\omega, \varepsilon_{it}^q, \varepsilon_{it}^y)$ are revealed to the firm after it makes it's quality choice and uncorrelated with the centers information set at the time of the quality, output choice, they do not impose an endogeneity problem. However, because the firms expectations about ω_{it}^h are a function of ω_{it}^q, we must still find a way of controlling for ω_{it}^h. From Proposition 1 we know that the firms expectation of productivity at the time of hiring can be recovered by inverting the firm's hiring policy at a fixed investment level,

$$\omega_{it}^h = h_{i_{it}}^{-1}(h_{it}, k_{it}, \ell_{it}, x_{it}). \tag{7}$$

Substituting (7) into (6), we arrive at our first-stage estimating equation,

$$\begin{aligned} y_{it} &= -\alpha_q q_{it} + \beta_k k_{it} + \beta_\ell \ell_{it} + h_{i_{it}}^{-1}(h_{it}, k_{it}, \ell_{it}, x_{it}) - \varepsilon_{it}^\omega - \alpha\varepsilon_{it}^q + \varepsilon_{it}^y. \\ &= -\alpha_q q_{it} + \Phi_{i_{it}}(h_{it}, k_{it}, \ell_{it}, x_{it}) + \varepsilon_{it}, \end{aligned} \tag{8}$$

where $\varepsilon_{it} = -\varepsilon_{it}^\omega - \alpha\varepsilon_{it}^q + \varepsilon_{it}^y$ and, $\Phi(h_{it}, k_{it}, \ell_{it}, x_{it}) = \beta_k k_{it} + \beta_\ell \ell_{it} + h^{-1}(h_{it}, k_{it}, \ell_{it}, x_{it})$. Due to invertibility requirements, we only have observations of (8) whenever hiring is non-zero.[16] Moreover, because the function $h_i^{-1}(\cdot)$ depends on the level of investment, we must estimate a separate $\Phi_i(\cdot)$ for each investment level. In practice, investment is zero xx percent of the time, we drop other investment levels and estimate (8) using observations where the center did not invest.[17] Finally, notice that the optimal policy for $q_{it} = q(k_{it}, \ell_{it}, x_{it}, \omega_{it}^q)$ whereas the optimal hiring policy is $h_{it} = h(k_{it}, \ell_{it}, x_{it}, \omega_{it}^h)$. Therefore, the difference between ω_{it}^q and ω_{it}^h provides variation needed to separately identify α_q.

Although the approach above handles the endogeneity of ω_{it}^q, we still have attenuation bias from the fact that ϵ_{it} and q_{it} are correlated through ε_{it}^q. To control for this attenuation bias,

[16] Because there are likely adjustment costs to hiring, $h_i^{-1}(\cdot)$ is not well defined when hiring is zero (multiple productivity levels may lead to zero net hiring). We follow the productivity literature and drop observations of zero hiring when estimating the first stage.

[17] I can do some experiments to show selection isn't a big deal here.

we use a second noisy measure of quality as an instrument in the second stage of a three stage estimation procedure following (Robinson 1988).[18] First we estimate $\widehat{E}[y|h_{it}, k_{it}, \ell_{it}, x_{it}, i_{it}]$ and $\widehat{E}[q|h_{it}, k_{it}, \ell_{it}, x_{it}, i_{it}]$ using standard nonparametric techniques, since the error term is uncorrelated with all of the regressors.[19] We then estimate $\hat{\alpha}_q$ with the linear instrumental variables regression,

$$y_{it} - \widehat{E}[y|h_{it}, k_{it}, \ell_{it}, x_{it}, i_{it}] = \alpha_q(q_{it} - \widehat{E}[q|h_{it}, k_{it}, \ell_{it}, x_{it}, i_{it}]) + \varepsilon_{it},$$

where we instrument for q_{it} with a second noisy measure of quality. In practice, we use the ratio of expected to actual deaths as this instrument, as discussed in Section 4.1. Finally, we can recover $\hat{\Phi}_i(\cdot)$ from the final nonparametric estimation,

$$y_{it} - \hat{\alpha}_q q_{it} = \Phi_{i_{it}}(h_{it}, k_{it}, \ell_{it}, x_{it}) + \varepsilon_{it}.$$

We recover the remaining parameters in a second stage. Note that, given any $\beta = (\beta_k, \beta_\ell)$, we can compute an estimate of unobserved productivity for each firm-year that has non-zero hiring,

$$\hat{\omega}_{it}(\beta) = \hat{\Phi}_{i_{it}}(h_{it}, k_{it}, \ell_{it}, x_{it}) - \beta_k k_{it} - \beta_\ell \ell_{it}.$$

Because ω_{it} follows a Markov process,

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}, \tag{9}$$

where g is a non-parametric function of ω_{it-1}, and ξ_{it} is a shock to productivity between time $t-1$ and t that is independent of the center's time-t information set.Thus, for any given $\beta = (\beta_k, \beta_\ell)$,

[18] An alternative approach, following Ackerberg et al. (2006), would have estimated y_{it} as a non-parametric function of $(q_{it}, h_{it}, k_{it}, \ell_{it}, x_{it}, i_{it})$ and then estimate α_q together with (β_k, β_ℓ) in the second stage. This would have the advantage of removing the requirement that q_{it} be flexibly chosen during the quality stage. However, the first stage estimation would be a nonparametric instrumental variables regression, introducing significant complications due to the high dimensionality of the problem.

[19] In practice, we approximate these functions using the method of sieves (Chen 2007).

we can estimate $g(\cdot)$ using the estimating equation:[20]

$$y_{it} + \hat{\alpha}_q q_{it} - \beta_k k_{it} - \beta_\ell \ell_{it} = g(\hat{\omega}_{it-1}(\beta)) + \eta_{it}(\beta),$$

which follows from substituting the production function from (6) into the innovation of productivity from (9), where $\hat{\alpha}_q$ is the consistent estimator of α_q recovered in the first stage. At the true value of β, $\eta_{it}(\beta) = \varepsilon_{it} + \xi_{it}$, and so, by construction, $\eta_{it}(\beta)$ is uncorrelated with the time-t labor and capital variables. Therefore, β can be consistently estimated using the moment conditions,

$$E\left[\begin{array}{c} \eta_{it}(\beta) k_{it} \\ \eta_{it}(\beta) \ell_{it} \end{array}\right] = 0. \tag{10}$$

We use (10) to estimate $\hat{\beta}$ via GMM, which can then be used to recover firm-level productivity estimates. Finally, standard errors are calculated using the block bootstrap, which accounts for statistical uncertainty in recovering the quality proxy, as well as both stages of the estimation process.

5 Results

NOTE: This section is currently under revision, these results should be considered preliminary.

5.1 Production Function Estimates and the Quality-Quantity Tradeoff

In order to compare our structural method to OLS and fixed-effects (FE) procedures, we present the results from estimates of dialysis centers' production functions using each technique in Table 4. For the results relating to the structural model, we use a fifth-order polynomial with interactions to approximate $\Phi(\cdot)$ in the first stage, and a fifth-order polynomial to approximate $g(\cdot)$ in the second stage.

The first three columns in Table 4 present results from specifications that do not include the infection rate as a proxy for quality, with the estimates of β_k and β_ℓ differing substantially

[20] We can estimate this equation using each observation that follows a non-zero hiring period. While it might seem more straightforward to recover $g(\cdot)$ by regressing $\hat{\omega}_{it}(\beta)$ on $\hat{\omega}_{it-1}(\beta)$, this would introduce a potential selection problem because we would only be able to use observations where hiring in period t itself was non-zero. We thank David Rivers for pointing this out to us.

Table 4: Model Estimates.

	Without Quality Choice			With Quality Choice		
	OLS	FE	Model	OLS	FE	Model
Expected Quality, α_q				0.0028	0.0018	0.0124
				(0.0007)	(0.0004)	(0.0042)
Capital, β_k	0.4628	0.1798	0.5210	0.4607	0.1788	0.5134
	(0.0208)	(0.0513)	(0.0446)	(0.0209)	(0.0514)	(0.0468)
Labor, β_ℓ	0.6709	0.1846	0.2527	0.6723	0.1855	0.2453
	(0.0149)	(0.0118)	(0.0304)	(0.0149)	(0.0119)	(0.0319)

across the three estimation methods. A comparison of our structural estimates in Column (III) to OLS in Column (I) and FE in Column (II) highlights several distinguishing features. First, OLS does not control for endogenous input choices, biasing the capital coefficient downwards and the labor coefficient upwards. This bias occurs because OLS relies on cross-sectional variation in stations to identify the labor and capital coefficients while ignoring the possibility of productivity differences across firms.

The FE procedure, in contrast, assumes productivity differences across firms remain constant over time and estimates the capital and labor coefficients on the basis of year-to-year changes in centers' inputs. Using this method, both the capital and labor coefficients fall substantially for two primary reasons. First, relying on only year-to-year variation makes measurement error in both capital and labor inputs a more prominent concern. Because stations and employees remain fairly stable over time, measurement error for hiring and investment decisions biases these coefficients towards zero.[21] A second potential reason for the discrepancy between the OLS and FE approaches is that capital and labor differences in the cross section may proxy for unobserved, time-invariant center characteristics (e.g., center size) that the FE specification captures through the productivity term. That the OLS results suggest centers have a production function with increasing returns to scale best illustrates this distinction, as we would expect that increasing the number of stations and staff within a center of constant size to exhibit decreasing returns to scale.

Finally, the third column presents results from estimates of the model presented in Section 3 with the added restriction that $\alpha_q = 0$. This specification employs a Markov process for

[21] For example, if a new station was installed in June of 2002, it will first be reported in 2003, but the difference in the number of patients served in 2002 versus 2003 will underreport the impact of the new station that actually came online for the second half of 2002.

productivity and uses both cross-section and time-series variation to identify the parameters, utilizing firms' hiring choices to identify unobserved productivity. These results exhibit decreasing returns to scale with respect to stations and staff, as expected. In addition, they indicate that the impact of additional stations is roughly twice that of increasing the number of staff, which seems natural given the production technology for dialysis procedures. While increasing the number of employees may allow a firm to treat more patients by speeding up the transition of a dialysis station from one patient to another, the number of patients being treated by the center at any given time is necessarily bounded by the number of available stations.

We next turn to the primary focus of the paper, estimating the quality-quantity tradeoff for dialysis centers, α_q. The final three columns of Table 4 present results from specifications that control for treatment quality using OLS, FE, and our structural model. All three specifications provide evidence of a statistically significant quantity-quality tradeoff, though the magnitude of the effect is much larger when using the structural model than with either the OLS or FE methods. The coefficient of 0.0124 from the structural model indicates that, holding inputs fixed, a firm that improves its quality enough so that its expected infection rate falls by 1 percentage point would need to reduce overall patient hours by 1.24 percent. Equivalently, a center could increase its output 1 percent by reducing quality such that its expected infection rate increases 0.80 percentage points. Alternatively, we can measure the cost of providing high-quality treatments in units of labor: a center can reduce its infection rate by 1 percent while maintaining its current level of output by increasing labor 5.1 percent. Given that the average center employs approximately 11 full-time-equivalent nurses, this roughly equates to hiring one additional part-time worker. Moreover, reducing the expected infection rate by a full standard deviation (6.3 percentage points) would cost the equivalent of roughly three full-time workers for the average center.

The smaller impact of quality on output in the OLS and FE specifications likely stems from endogeneity bias. We would expect, and in fact verify below in Table 8, that providing high-quality treatments is positively associated with productivity. Since the OLS specification does not control for differences in productivity, an estimate of α_q in this setup will be biased downward (recall that α_q enters the estimating equation (8) with a negative sign). While the FE approach controls for time-invariant productivity levels, if firms' changes in quality targets are positively correlated with changes in their productivity, the FE estimate of α_q will also be biased downwards. This effect, coupled with the effects of attenuation bias already discussed

Table 5: Robustness Checks.

	I	II	II	IV
Quality Effort, α_q	-0.0124	-0.0106	-0.0101	-0.0121
	(0.0042)	(0.0036)	(0.0042)	(0.0009)
Capital, β_k	0.5134	0.5077	0.4381	0.5136
	(0.0468)	(0.0474)	(0.0553)	(0.0466)
Labor, β_ℓ	0.2453	0.2448	0.1989	0.2455
	(0.0319)	(0.0313)	(0.0169)	(0.0318)
Control for Patient Characteristics	YES	NO	YES	YES
Control for Market Characteristics	YES	YES	NO	YES
Instrument for Quality	YES	YES	YES	NO

above, biases the estimates of the quality-quantity tradeoff towards zero.

In Table 5, we consider several robustness checks of the baseline results, which are repeated in the first column. The second column drops controls for patient characteristics and instead simply uses the infection rate itself as a proxy for quality targets. The third column drops the center characteristics of for-profit status and competition from the hiring function. Finally, the fourth column does not instrument for the quality proxy but instead simply uses OLS to estimate the first stage. In all cases, the effect of quality declines slightly, though our estimate of a significant quality-quantity tradeoff remains robust to various model specifications.

5.2 Productivity Dispersion, Growth, and Persistance

Having estimated the firm-level production function, we are able to recover center-year (log) productivity from

$$\hat{\omega}_{it} = y_{it} - \hat{\alpha}_0 + \hat{\alpha}_q q_{it} - \hat{\beta}_k k_{it} - \hat{\beta}_\ell \ell_{it};$$

this allows us to analyze the dispersion, growth, and persistence of productivity within the dialysis industry. Moreover, we are able to estimate the importance of productivity for firms' quality choices.

To asses the extent of productivity dispersion, we first calculate the proportion of the variance in output explained by the production function outside of productivity differences:

$$R^2 = 1 - \frac{V(\hat{\omega}_{it})}{V(y_{it})}.$$

Our results indicate that the amount of productivity dispersion in the dialysis industry is substantial, with $R^2 = 0.489$, meaning that about half of the variation in output is attributable to productivity differences across firms, not input or quality differences. For a basis of comparison, Fox & Smeets (2011) report R^2 statistics for service industries ranging from 0.438 (Accounting) to 0.739 (Computer Activities).

We can then use these productivity estimates to measure productivity growth and persistence within the dialysis industry, as reported in Table 6. Overall, average productivity for the industry is roughly constant over the sample, with a slight drop in 2008. Again, we find significant productivity dispersion across the industry: the inter-quartile range indicates that a firm at the 75th percentile of productivity is over 50 percent more productive than one at the 25th percentile.

On average, productivity growth at the firm level is extensive, ranging between 4 and 7 percent per year; at the same time, we observe a large degree of variation in productivity growth within the sample. The contrast of large firm-level productivity growth with slow industry-wide productivity growth suggests that firms enter at a lower level of productivity than incumbents.[22] But despite the high average growth rates, substantial dispersion in growth rates across centers still remains, suggesting that year-to-year productivity shocks have a substantial impact on centers' output. These shocks could result from high staff turnover, changes in patient characteristics, or other factors that affect productivity. We also find that productivity is persistent within a firm across years, as shown by a correlation in log productivity of approximately 0.8 for the entire sample.

We further explore the trends in productivity across firms in Table 7. Here, we stratify centers by age, determined by the year in which they first appear in the sample.[23] The average productivity of firms increases substantially with age, while the dispersion in productivity falls with each age group. Note, however, that the increase in productivity from age 0 to 1 is at least partially due to centers only operating for a portion of their initial year, and the results for productivity growth in years 1-3 indicate a fast but declining rate of productivity growth over the initial years of a center's existence. In contrast, the average productivity growth rate of 1.5 percent indicates only modest growth for established firms. Overall, it appears that new

[22]The decline in the number firms in 2008 is due to incomplete reporting of centers' staffing levels rather than actual closures, which are rare in this industry. For the purposes of estimation, we assume these data are missing at random.

[23]Centers appearing in 2004 are assumed to be four or more years old.

Table 6: (Log) Productivity, Productivity Growth, and Persistence.

Year	N	Mean Level	St. Dev. Level	IQR Level	Mean Gain	St. Dev. Gain	Corr($\omega_{it}, \omega_{i,t-1}$)
2004	3,360	1.6646	0.5350	0.5043			
2005	3,563	1.6667	0.5279	0.5128	0.0682	0.3081	0.8191
2006	3,733	1.6783	0.4879	0.5151	0.0511	0.3113	0.7712
2007	3,885	1.6615	0.4979	0.5121	0.0385	0.2843	0.8130
2008	3,754	1.6177	0.5283	0.5253	0.0406	0.3020	0.7935
Total	18,295	1.6575	0.5155	0.5151	0.0492	0.3015	0.7995

Notes: Mean level is the average log productivity of all centers active in year t. Mean gain is the average change in log productivity of centers active in years t and $t-1$.

Table 7: (Log) Productivity Growth of New Firms.

Age (years)	N	Mean Level	St. Dev. Level	IQR Level	Mean Gain	St. Dev. Gain	Corr($\omega_{it}, \omega_{i,t-1}$)
0	910	0.6858	1.0115	1.3557			
1	668	1.4528	0.4673	0.6018	0.6652	0.7363	0.6947
2	447	1.5633	0.4098	0.5309	0.1012	0.1985	0.9000
3	232	1.6041	0.3973	0.4970	0.0450	0.1968	0.8812
4+	16,038	1.7246	0.4111	0.4644	0.0147	0.2199	0.8475
Total	18,295	1.6575	0.5155	0.5151	0.0492	0.3015	0.7995

firms enter with productivity levels well below the industry average, but then experience strong growth to "catch up" to established firms. This, coupled with the slow growth of established firms themselves, results in relatively stagnant productivity growth for the industry as a whole.

5.3 The Determinants of Quality

The quality-quantity tradeoff estimated in Section 5.1 highlights the costs of providing high-quality care. In this section, we examine the firm-level characteristics associated with providing better care. In particular, we find that non-profit firms, which may have objectives beyond maximizing profits, tend to provide higher-quality treatments.

Quality provision within our model is a non-parametric function of all the factors that affect productivity, the for-profit status of the firm, and the extent of competition the firm faces. To summarize the relationship between quality and its determinants, we estimate,

Table 8: Partially Linear Quality Regressions.

	I	II	III	IV	V
For Profit	-1.5603		-1.5390	-1.5444	
	(0.2021)		(0.2030)	(0.2111)	
Monopolist		0.5390	0.4824		0.4725
		(0.2211)	(0.2196)		(0.2222)
Duopolist		-0.2474	-0.2977		-0.2926
		(0.1876)	(0.1843)		(0.1855)
Triopolist		-0.4701	-0.4678		-0.4431
		(0.2257)	(0.2234)		(0.2224)
Nonparametric Control for:					
Productivity	Yes	Yes	Yes	Yes	Yes
Capital	Yes	Yes	Yes	Yes	Yes
Labor	Yes	Yes	Yes	Yes	Yes
For-Profit Status	No	No	No	No	Yes
Competition	No	No	No	Yes	No

$$q_{it} = \delta_{fp(it)} + \gamma_{c(it)} + \mu(k_{it}, \ell_{it}, \hat{\omega}_{it}) + \zeta_{it}. \tag{11}$$

As before, q_{it} is our proxy for quality, which is the deviation in the estimated infection rate from expectations. We use a fifth-order polynomial sieve to approximate the nonparametric function, μ, which contains the variables affecting the firm's production possibilities frontier; to proxy for productivity, we use our productivity estimate recovered from the production function estimation. The parameters of interest are then $\delta_{fp(it)}$, a dummy coefficient for whether the firm is for-profit, and $\gamma_{c(it)}$, which is a set of dummy coefficients representing the extent of competition faced by the center in its local market (0, 1, 2, or 3 or more competitors). While this regression suffers from specification and measurement error (both q_{it} and $\hat{\omega}_{it}$ are contaminated with measurement error), it remains indicative of centers' quality policies. Again, we compute standard errors using the block bootstrap, which incorporates statistical sampling uncertainty in estimates of quality and productivity and controls for firm-level serial correlation.[24]

We present several versions of this quality regression in Table 8. In Column I, we examine the effect of for-profit status on quality while using non-parametric controls for labor, quality, and productivity. The results show that for-profit firms provide significantly worse care than non-profit firms, with the expected infection rate more than 1.5 percentage points (over 10 percent)

[24]We have also estimated several linear specifications, which yield similar results.

higher at for-profits. This estimate is statistically significant and provides strong evidence that firms respond to profit-based incentives by delivering lower-quality treatments. Not only is it costly to provide high-quality care, firms with stronger incentives to control costs appear to respond by delivering lower-quality treatments. Therefore, policymakers aiming to incentivize healthcare providers to reduce costs must fully consider the implications of their initiatives for patient outcomes.

Column II examines the impact of competition on quality, where the base category is centers with three or more competitors in their market (defined as an HSA). The results show no clear pattern between competition and firms' quality choices. Surprisingly, monopolists tend to provide *higher*-quality treatments than firms in more competitive markets, counter to the intuition that competition for patients might provide an incentive to improve quality. On the other hand, duopolists and triopolists offer weakly lower quality than firms with three or more competitors. Overall, competition does not appear to provide a strong incentive for firms to improve their quality of care.

The results of the first two columns are robust to several other specifications of the quality policy function. In Column III, we include both the for-profit and the competition dummies together. Column IV estimates the for-profit dummy while allowing for nonparametric controls of competition by estimating a separate μ for each competition status. Finally, Column V estimates the competition dummies while controlling for for-profit status non-parametrically. These various specifications confirm our original findings. Controlling for productivity differences, for-profits offer substantially lower-quality treatments than for-profits, and are therefore able to treat more patients using the same amount of resources. By contrast, the extent of competition in a market does not appear to influence the quality-quantity choices of firms in a systematic way.

Because non-profit centers may face less urgency to maximize profits than for-profit centers, for-profit centers likely choose to treat more patients by providing lower-quality care given the substantial quality-quantity tradeoff we have documented for dialysis treatments. And while competition might lead centers to improve their quality of care in order to attract more patients, our results do not provide strong evidence that this is the case. Because demand for dialysis treatments is relatively inelastic — most patients simply choose the center closest to their homes — the limited influence of competitive forces on patient outcomes may not be surprising, though it does suggest that policies aimed at improving the quality of care through fostering greater

patient choice in the dialysis industry may be ineffective.

6 Conclusion

By estimating center-level production functions that incorporate endogenous quality choices, we find evidence that dialysis centers face a tradeoff between treating more patients and providing higher-quality care. Our findings suggest that policies aimed at increasing efficiency may inadvertently affect health outcomes. Although we find considerable dispersion in productivity across firms, these results imply that incentives to reduce costs may lead to lower-quality care, not greater efficiency. Similarly, our results for non-profit centers also provide evidence that firms react to cost incentives by adjusting the quality of their treatments. Non-profit centers, which presumably face less urgency to reduce costs, provide higher-quality care than their for-profit counterparts.

We find little evidence that market forces discipline centers to provide high-quality care. While competition might be expected to provide a demand-side incentive for improving quality, we find that firms in more-competitive markets are not more likely to offer better care than monopolists. Disentangling the potential explanations for this result lies beyond this paper, though the inelastic demand for dialysis treatments, the dominance of two for-profit chains, and the weak incentives imposed by Medicare all likely contribute to this outcome.

Because dialysis treatments comprise a large — and growing — expense for Medicare, controlling the cost of dialysis provision will likely concern policy makers for the foreseeable future. Our work informs these policy discussions by showing that, while productivity dispersion is extensive within the industry, cost-cutting initiatives may simply reduce the quality of care provided rather than promote efficiency. More importantly, because dialysis resembles other healthcare settings, these findings illustrate the challenges of introducing policies intended to minimize costs while maintaining high standards of care.

References

Ackerberg, D. A., Caves, K. & Frazer, G. (2006), Structural identification of production function. UCLA, Deloitte and Touche, Rotman School of Management.

Blundell, R. & Bond, S. (2000), 'Gmm estimation with persistent panel data: an application to production functions', *Econometric Reviews* **19**, 321–340.

Chen, X. (2007), Large sample sieve estimation of semi-nonparametric models, *in* J. J. Heckman & E. E. Leamer, eds, 'Handbook of Econometrics', Vol. 6B, Elsevier Science B.V., Amsterdam, chapter 74, pp. 5549–5632.

Dai, M. (2012), Specialization and differentiation in outpatient dialysis. Drexel Lebow.

De Loecker, J. (2011), 'Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity', *Econometrica* **79**(5), 1407–1451.

Ford, J. M. & Kaserman, D. L. (2000), 'Ownership structure and the quality of medical care: evidence from the dialysis industry', *Journal of Economic Behavior & Organization* **43**, 279–293.

Fox, J. & Smeets, V. (2011), 'Does input quality drive measured productivity differences in firm productivity', *International Economic Review* **52**(4), 961–989.

Gandhi, A., Navarro, S. & Rivers, D. (2011), On the identification of production functions: How heterogeneous is productivity? University of Wisconsin-Madison and University of Western Ontario.

Imerman, M. & Otto, D. (2004), Preliminary Market and Cost Analysis of a Five-Station Hemodialysis Facility in Marengo, Iowa. Iowa State University.

Lee, J., McCullough, J. S. & Town, R. J. (2012), The impact of health information technology on hospital productivity. National Bureau of Economic Research.

Levinsohn, J. & Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', *The Review of Economic Studies* **70**(2), pp. 317–341.
URL: *http://www.jstor.org/stable/3648636*

Mundalk, Y. (1961), 'Empirical production function free of management bias', *Journal of Farm Economics* **43**, 44–56.

Olley, G. S. & Pakes, A. (1996), 'The dynamics of productivity in the telecommunications equipment industry', *Econometrica* **64**(6), pp. 1263–1297.
URL: *http://www.jstor.org/stable/2171831*

Pakes, A. (1994), The estimation of dynamic structural models: Problems and prospects, *in* J. J. Laffont & C. Sims, eds, 'Advances in Econometrics: Proceedings of the 6th World Congress of the Econometric Society', Vol. II, pp. 171–259.

Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., Sexton, B., Hyzy, R., Welsh, R., Roth, G. et al. (2006), 'An intervention to decrease catheter-related bloodstream infections in the icu', *New England Journal of Medicine* **355**(26), 2725–2732.

ProPublica (2011), 'Dialysis: High costs and hidden perils of a treatment guaranteed to all'.
URL: *http://www.propublica.org/series/dialysis*

Ramanarayanan, S. & Snyder, J. (2011), Reputations and firm performance: Evidence from the dialysis industry. UCLA Anderson.

Robinson, P. (1988), 'Root-n-consistent semiparametric regression', *Econometrica* **56**(4), 931–954.

Romley, J. A. & Goldman, D. P. (2011), 'How costly is hospital quality? a revealed-preference approach', *The Journal of Industrial Economics* **59**(4), 578–608.
URL: *http://dx.doi.org/10.1111/j.1467-6451.2011.00468.x*

Sloan, F. (2000), Not-for-profit ownership and hospital behavior, *in* A. Culyer & J. Newhouse, eds, 'Handbook of Health Economics', Vol. 1B, Elsevier Science B.V., Amsterdam.

Syverson, C. (2011), 'What determines productivity?', *Journal of Economic Literature* **49**(2), 326–65.

USRDS (2010), 2010 Annual Data Report, Technical report, United States Renal Data System, Minneapolis, MN.

A Proofs

Proof of Lemma 1 *The center's expected period-return to labor is increasing in ω^q. I.e., $\frac{\partial \pi}{\partial \ell}$ is increasing in ω^q.*

Proof. Since center payoffs are increasing both y and q (i.e., the center has non-satiable payoffs), we know that the center will choose $(\tilde{y}, \tilde{q})$ to solve the following problem where the production constraint binds,

$$\begin{aligned} \pi(k, \ell, x, \omega^q) = \max_{\tilde{y}, \tilde{q}} \; & E[\rho(y, q, k, \ell, x)] \\ \text{subject to: } & T(\tilde{y}, \tilde{q}) = F(k, \ell, \omega^q) \\ & y = \tilde{y} + \varepsilon^y \\ & q = \tilde{q} + \varepsilon^q \end{aligned}$$

Totally differentiating π with respect to ℓ, the return to an increase in labor is,

$$\frac{d\pi}{d\ell} = E\left[\rho_y \frac{d\tilde{y}}{d\ell} + \rho_q \frac{d\tilde{q}}{d\ell} + \rho_\ell\right],$$

where ρ_x represents the partial derivative of ρ with respect to x and the total derivatives with respect to $\tilde{y}$ and $\tilde{q}$ are the center's optimal policy change for a change in ℓ. We know both are weakly positive, with at least one strictly positive, because an increase in ℓ relaxes the production constraint through an increase in $F(\cdot)$, and $\rho(\cdot)$ is increasing in both y and q. To see that this is increasing in ω^q, note that an increase in ω^q also relaxes the production constraint. Differentiating again with respect to ω^q yields,

$$\frac{d^2\pi}{d\ell d\omega^q} = E\left[\rho_y \frac{d\tilde{y}}{d\ell}\frac{d\tilde{y}}{d\omega^q} + \rho_q \frac{d\tilde{q}}{d\ell}\frac{d\tilde{q}}{d\omega^q}\right].$$

Non-satiation again ensures that both terms are weakly positive and at least one is strictly positive. □

Proof of Proposition 1 *The for any fixed investment level κ, firm hiring function $h(k, \ell, x, \omega^h)$ is invertible with respect to ω^h on the domain $\{(k, \ell, x, \omega^h) : i(k, \ell, x, \omega^h) = \iota\}$,*

$$\omega^h = h_\iota^{-1}(k, \ell, x, h).$$

Proof. We will apply (Pakes 1994, Theorem 1) accounting for three differences which complicate our model. First, following (Pakes 1994, Lemma 1), we note the the inclusion of a discrete choice of capital investment does not alter our ability to use the firm's first order condition with respect to hiring, we must simply substitute the (observed) optimal investment choice ι into the first order condition,

$$c_h(\iota, h) + \beta E V_h(k + \iota, \ell + h, x', \omega^{q'}) | k, \ell, \omega^h, \iota, h] = 0.$$

Second, because x evolves according to an exogenous stochastic process, we can use the insights of (De Loecker 2011, Appendix C) that an additional exogenous variables do not alter the invertibility property. The only remaining difference between this problem and the traditional investment problem described by Olley & Pakes (1996) is that our productivity process evolves intra-period between the quality and investment stages. However, because $Pr(\omega^{q'}|\omega^h)$ and $Pr(\omega^h|\omega^q)$ are both stochastically increasing in ω^h and ω^q respectively. (The former by assumption, and the latter because it is a random walk.) We know that $Pr(\omega^{h'}|\omega^h)$ is also

stochastically increasing. We can now write a single Bellman equation for a center at the time of the hiring decision as,

$$V(k,\ell,x,\omega^q,\xi,\omega^h) = \max_{i,h} -c(i,h,k,\ell)+\pi(k,\ell,x,\omega^q)+\xi+\beta E[V(k',\ell',x',\omega^{q'},\xi,\omega^{h'})|k,\ell,x,\omega^h,i,h]$$

Today's realized profits from the quality stage are $\pi(k,\ell,x,\omega^q)+\xi$, where ξ is uncorrelated with the agent's information set at the time of the quality choice (or any time before the quality choice), but is known at the time of the hiring decision since production outcomes are already revealed. They are sunk with respect to today's hiring decision. Note that ω^q and ξ represent two additional state variables, but they both evolve exogenously. Moreover, conditional on ω^h they are uncorrelated with future draws of ω^q and ξ, which is why they do not appear in the final expectation term. Finally, using Lemma 1 and the fact that $Pr(\omega^{q'}|\omega^h)$ is stochastically increasing, we know $E[\frac{\partial\pi(k',\ell',x'\omega^{q'})}{\partial\ell}|k,\ell,x,\omega^h]$ is increasing in ω^h.

Following De Loecker (2011), group $k^* = (k,\ell,x,\omega^q,\xi)$. So the policy function can be written $h(k^*,\omega^h)$. We can now directly apply Pakes (1994, Lemma 3) where $c(h,\iota,k^*)$ stands for $c(x,k)$ (recall ι is the optimal capital investment decision), $\pi(\omega^q,k^*) = \pi(k,\ell,x,\omega^q)+\xi$ for $\pi(\omega,k)$ and the choice variable is h (hiring), rather than x in (continuous capital investment in Pakes 1994). □